AD/A-003 893

SOME REMARKS ON THE FINITE-MEMORY K-HYPOTHESES PROBLEMS

Bruno O. Shubert

Naval Postgraduate School

Prepared for:

Office of Naval Research

October 1974

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

The work reported herein was supported by the Office of Naval Research through the Research Foundation Funds of the Naval Postgraduate School.

This report was prepared by:

Bruno Shubert
Bruno O. Shubert
Associate Professor

Reviewed by:

David A. Schrady
David A. Schrady, Chairman
Department of Operations Research
and Administrative Sciences

David B. Hoisington
Acting Dean of Research

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER
NPS55Sy74101

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE *(and Subtitle)*
Some Remarks on the Finite-Memory
K-Hypotheses Problems

5. TYPE OF REPORT & PERIOD COVERED
Technical Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Bruno O. Shubert

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research, Arlington, VA 22217

12. REPORT DATE
October 1974

13. NUMBER OF PAGES
~~28~~ 27

14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*

15. SECURITY CLASS. *(of this report)*
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
US Department of Commerce
Springfield, VA. 22151

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*
Finite memory, Markov chains

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*
Finite-memory statistical problems typically deal with the situation where the class of statistics is restricted to those taking on a fixed finite number of values. Although a potentially infinite number of samples may be available the statistician is allowed to base his inference only on the current value of such a statistic - the current state of his finite memory. This is the case for instance when the inference is to be performed by a small size computer.

During the past several years a number of results have been obtained concerning a two-hypotheses finite-memory problem. In this report we consider

(27)

20. some aspects of the case where the number of hypotheses is greater than two. In particular we derive a bound on the error probability for the 3-hypothesis case, present a counterexample to a recently proposed conjecture and briefly discuss a finite-memory version of the minimax theorem. We also include two appendices containing some results on finite Markov chains.

1. INTRODUCTION

Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables taking on values in some measurable space $\mathcal{X}$. Let $P_1, \ldots, P_K$ be a finite collection of probability measures on $\mathcal{X}$, and let $H_k, k = 1, \ldots, K$, denote the hypothesis that the common distribution of the X_n's is P_k.

We wish to associate with the observed sequence $X_1, X_2, \ldots$ a sequence of decisions $d_1, d_2, \ldots$, $d_n \in \{H_1, \ldots, H_K\}$ about the true hypothesis H. However, the decision d_n at time n is allowed to depend on $X_1, \ldots, X_n$ only through a finite-valued statistic $T_n \in \{1, \ldots, m\}$, which represents the current state of the memory. This statistic is updated after each observation, i.e.,

$$T_{n+1} = f(T_n, X_n) \ , \quad n=1,2,\ldots \ ,$$

where $f : \{1,\ldots,m\} \times \mathcal{X} \to \{1,\ldots,m\}$ is a time-invariant updating rule. The decision d_n is then given by

$$d_n = d(T_n) \ , \quad n=1,2,\ldots \ ,$$

where $d : \{1,\ldots,m\} \to \{H_1,\ldots,H_K\}$ is a time-invariant decision rule. Let for a given f and d

$$P_e^{(k)}(f,d) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P(d_n \neq H_k) \tag{1.1}$$

be the asymptotic expected frequency of incorrect decisions if the true hypothesis is H_k. Our goal is to find (f,d) which minimizes

$$P_e(f,d) = \max_{k=1,\ldots,K} P_e^{(k)}(f,d) \tag{1.2}$$

(Thus we have chosen the minimax criterion. An alternate approach would be to minimize

$$P_e(\underline{\pi};(f,d)) = \sum_{k=1}^{K} \pi_k P^{(k)}(f,d) ,$$

where $\pi = (\pi_1,\dots,\pi_K)$ is a prior distribution on $\{H_1,\dots,H_K\}$. In this report we consider the former.)

The pair (f,d) together with the domains and ranges of the two functions is formally equivalent to the definition of a finite automaton (see e.g. [1]). The automaton has $S = \{1,\dots,m\}$ as its state space, X as its input space, $\{H_1,\dots,H_K\}$ as its output space, and f and d as its state-transition function and state-output function respectively. If the sequence $X_1,X_2,\dots$ of i.i.d. random variables is applied to the input of such an automaton the resulting sequence of states $T_1,T_2,\dots$ is then a time-homogeneous Markov chain with transition probabilities

$$p_{ij} = P_k(\{x\in X : f(i,x) = j\}) , \quad i,j\in S . \tag{1.3}$$

Hence the limit in (1.1) always exists. If the state-transition function f is such that the resulting chain is regular then in fact

$$P_e^{(k)}(f,d) = \mu_k(d^{-1}(H_k)) ,$$

where μ_k is the stationary distribution on S . Throughout this paper we assume that this is the case, i.e., we consider only transition functions which yield regular Markov chains under each hypothesis.

Following Hellman and Cover [3] we would like to include the possibility that the transition function f can be randomized. One way of defining such a randomization would be to introduce another input sequence $Y_1,Y_2,\dots$ of i.i.d. random variables, independent of the sequence $X_1, X_2, \ldots,$ and uniformly distributed on the interval $[0,1]$. The transition probabilities (1.3) would then be

$$p_{ij} = E_k\{p_{ij}(X)\} , \quad \text{where} \tag{1.4}$$

$$p_{ij}(x) = \lambda(\{y \in [0,1] : f(x,y,i) = j\}) , \tag{1.5}$$

λ being Lebesgue measure on $[0,1]$.

However, we find it more convenient to express the randomized state transition function f as a pair (A, Δ) as follows:

$$A = \{A_{ij} : i=1,\ldots,m \; ; \; j=1,\ldots,m \; ; \; i \neq j\} ,$$

where A_{ij} are measurable subsets of X .

$$\Delta = \{\delta_{ij} : i=1,\ldots,m \; ; \; j=1,\ldots,m \; ; \; i \neq j\} ,$$

where $\delta_{ij} \geq 0$ and $\sum_j \delta_{ij} \leq 1$ for all i,j .

The transition probabilities (1.5) if $X = x$ is observed are now defined by

$$p_{ij}(x) = \delta_{ij} \quad \text{whenever} \quad A_{ij} \ni x \quad \text{for} \quad i \neq j ,$$

$$p_{ii}(x) = 1 - \sum_{j \neq i} p_{ij}(x) ,$$

and (1.3) becomes

$$\left. \begin{aligned} p_{ij} &= P_k(A_{ij})\delta_{ij} \quad \text{if} \quad i \neq j , \\ p_{ii} &= 1 - \sum_{j \neq i} p_{ij} : \end{aligned} \right\} \tag{1.6}$$

We will refer to the triplet (A, Δ, d) as <u>randomized finite automaton</u> (RFA) and to the set Δ as <u>randomization</u>.

Notice the class of all randomization is closed with respect to multiplication of corresponding elements, that is if $\Delta = \{\delta_{ij}\}$ and $\Delta' = \{\delta'_{ij}\}$ are randomization then $\Delta\Delta' = \{\delta_{ij}\delta'_{ij}\}$ is again a randomization. Notice also that the sets A_{ij} need not be disjoint.

We now present a simple lemma to be used in the remaining sections.

Lemma 1: Let (A, Δ, d) be a RFA, let μ_k , $k = 1,\ldots,K$, be stationary distributions of the resulting Markov chain of states. Let $R = [r_{k\ell}]$ be a $K \times K$ matrix with positive entries

$$r_{k\ell} = \frac{\mu_k(d^{-1}(H_\ell))}{\mu_k(d^{-1}(H_k))} ,$$

let $\rho(A, \Delta, d)$ be the maximal eigenvalue of R .

Then

$$P_e(A, \Delta, d) \geq 1 - \frac{1}{\rho(A, \Delta, d)} ,$$

and there exists a randomization Δ' such that

$$P_e(A, \Delta\Delta', d) = 1 - \frac{1}{\rho(A, \Delta, d)} .$$

Proof:

$$(1-P_e)^{-1} = (1-\max_k P_e^{(k)})^{-1}$$

$$= \max_k (1-P_e^{(k)})^{-1} = \max_k (\mu_k(d^{-1}(H_k)))^{-1}$$

$$= \max_k \sum_{\ell=1}^{K} r_{k\ell} \geq \rho ,$$

since by Perron-Frobenius theorem the maximal eigenvalue of a positive matrix can never exceed the largest of the row-sums.

To prove the second statement let $\underline{v} = (v_1, \dots, v_K)$ be an eigenvector corresponding to $\rho(A, \Delta, d)$ and normalized such that

$$v_k > 0 \;, \quad k = 1, \dots, K \;, \quad v_1 + v_2 + \cdots + v_K = a \;,$$

where $0 < a \le 1$ is an arbitrary constant. (This is always possible since the matrix R is positive.) Define the randomization $\Delta' = \{\delta'_{ij}\}$ by

$$\delta_{ij} = \frac{1}{u_i} \;, \quad j \neq i \;, \quad \text{where} \quad u_i = v_k \quad \text{whenever} \quad i \in d^{-1}(H_k) \;.$$

Let $p_{ij}^{(k)}$ and $p_{ij}^{\prime (k)}$ be transition probabilities and μ_k and μ'_k the stationary distribution of (A, Δ, d) and $(A, \Delta\Delta', d)$ respectively. Then by (1.6) for any k and $i \neq j$

$$p_{ij}^{\prime (k)} = \frac{1}{u_i} p_{ij}^{(k)} \;,$$

and hence for any partition of the state space S into two subsets S_1 and S_2 we must have

$$\sum_{i \in S_1} \sum_{j \in S_2} p_{ij}^{(k)} \mu_k(i) = \sum_{i \in S_2} \sum_{j \in S_1} p_{ij}^{(k)} \mu_k(i)$$

$$\sum_{i \in S_1} \sum_{j \in S_2} \frac{p_{ij}^{(k)}}{u_i} \mu'_k(i) = \sum_{i \in S_2} \sum_{j \in S_1} \frac{p_{ij}^{(k)}}{u_i} \mu'_k(i) \;.$$

Thus $\mu'_k(i) = C_k u_i \mu_k(i)$ for all i, k with $C_k > 0$ independent of i so that

$$\mu'_k(d^{-1}(H_\ell)) = C_k v_\ell \mu_k(d^{-1}(H_\ell))$$

for all $k, \ell = 1, \dots, K$. But then for all k

$$\sum_{\ell=1}^{K} r'_{k\ell} = \sum_{\ell=1}^{K} r_{k\ell} \frac{v_\ell}{v_k} = \rho(A,\Delta,d)$$

since $\underline{v}$ is an eigenvector of $\rho(A,\Delta,d)$.

Q.E.D.

2. UPPER BOUND ON THE ERROR PROBABILITY FOR THE 3-HYPOTHESES 3-STATE PROBLEM

Let $K = 3$, $m = 3$, and let

$$P_e^* = \inf P_e(A,\Delta,d) ,$$

the infimum being taken over the class of all 3-state RFA. Let for $i,j = 1,2,3$

$$\gamma_{ij} = \frac{\sup\limits_{A\subset X} P_i(A)/P_j(A)}{\inf\limits_{A\subset X} P_i(A)/P_j(A)} \tag{2.1}$$

let

$$\bar{g} = (1/3)(\gamma_{12}^{-1} + \gamma_{23}^{-1} + \gamma_{13}^{-1}) ,$$

$$\bar{G} = (1/2)\max\{\gamma_{12}^{-1} + \gamma_{23}^{-1} , \gamma_{12}^{-1} + \gamma_{13}^{-1} , \gamma_{23}^{-1} + \gamma_{13}^{-1}\} .$$

In this section we show that

$$P_e^* \le 1 - (1 + 2\bar{g}^{1/2} \cosh(1/3 \operatorname{argcosh} \bar{G}\, \bar{g}^{-3/2}))^{-1} . \tag{2.2}$$

We also establish a simpler but looser bound, namely

$$P_e^* \le 1 - (1 + 2\bar{G}^{1/3})^{-1} . \tag{2.3}$$

Notice that (2.3) implies that if $\gamma_{12} = \gamma_{23} = \gamma_{13} = +\infty$ then $P_e^* = 0$, a result obtained by Sagalowicz in [4] and extended later by Yakowitz in [5].

Proof of (2.2) and (2.3):

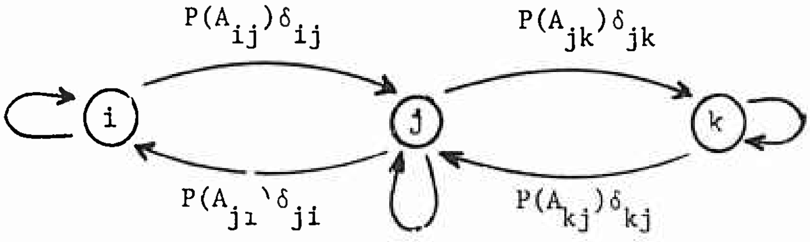
Let the letters i,j,k stand for a permutation of 1,2,3, let $\varepsilon > 0$. Let $A_{ij}(\varepsilon)$ be measurable subsets of X such that

$$\frac{P_i(A_{ij}(\varepsilon))}{P_j(A_{ij}(\varepsilon))} \leq \inf_A \frac{P_i(A)}{P_j(A)} + \varepsilon ,$$

let

$$\gamma_{ij}(\varepsilon) = \frac{P_j(A_{ij}(\varepsilon))}{P_i(A_{ij}(\varepsilon))} \frac{P_i(A_{ji}(\varepsilon))}{P_j(A_{ji}(\varepsilon))}$$

so that $\gamma_{ij}(\varepsilon) \rightarrow \gamma_{ij}$ as $\varepsilon \rightarrow 0$. Consider a 3-state RFA $(\mathcal{A},\Delta ,d)$, where $\mathcal{A} = \{A_{ij}(\varepsilon)\}$, $\Delta = \{\delta_{ij}\}$ such that $\delta_{ik} = \delta_{ki} = 0$, otherwise arbitrary, and $d(i) = H_i$, $d(j) = H_j$, $d(k) = H_k$.



The stationary distribution of the resulting Markov chain of states are (see Appendix A)

$$\mu(i) = cP(A_{ji})\delta_{ji}P(A_{kj})\delta_{kj} ,$$

$$\mu(j) = cP(A_{ij})\delta_{ij}P(A_{kj})\delta_{kj} ,$$

$$\mu(k) = cP(A_{jk})\delta_{jk}P(A_{ij})\delta_{ij} ,$$

where c is a normalizing constant. (The epsilon has been dropped temporarily to ease the notation.) The matrix R of Lemma 1 is given by

$$R = \begin{bmatrix} 1 & , \dfrac{P_i(A_{ij})\delta_{ij}}{P_i(A_{ji})\delta_{ji}} & , \dfrac{P_i(A_{jk})P_i(A_{ii})\delta_{jk}\delta_{ij}}{P_i(A_{ji})P_i(A_{kj})\delta_{ji}\delta_{kj}} \\ \dfrac{P_j(A_{ji})\delta_{ji}}{P_j(A_{ij})\delta_{ij}} & , \quad 1 & , \dfrac{P_j(A_{jk})\delta_{jk}}{P_j(A_{kj})\delta_{kj}} \\ \dfrac{P_k(A_{ji})P_k(A_{kj})\delta_{ji}\delta_{kj}}{P_k(A_{jk})P_k(A_{ij})\delta_{jk}\delta_{ij}} & , \dfrac{P_k(A_{kj})\delta_{kj}}{P_k(A_{jk})\delta_{jk}} & , \quad 1 \end{bmatrix} ,$$

and its characteristic equation has the form

$$(1-\lambda)^3 - (1-\lambda)C_\varepsilon + D_\varepsilon = 0 \quad , \tag{2.4}$$

where

$$C_\varepsilon = \gamma_{ij}^{-1}(\varepsilon) + \gamma_{jk}^{-1}(\varepsilon) + \frac{P_i(A_{jk})P_i(A_{ij})P_k(A_{ji})P_k(A_{kj})}{P_i(A_{ji})P_i(A_{kj})P_k(A_{jk})P_k(A_{ij})}$$

and

$$D_\varepsilon = \frac{P_i(A_{ij})P_j(A_{jk})P_k(A_{ji})P_k(A_{kj})}{P_i(A_{ji})P_j(A_{kj})P_k(A_{jk})P_k(A_{ij})}$$

$$+ \frac{P_j(A_{ji})P_k(A_{kj})P_i(A_{jk})P_i(A_{ij})}{P_j(A_{ij})P_k(A_{jk})P_i(A_{ji})P_i(A_{kj})} \quad .$$

Now D_ε can be written as

$$D_\varepsilon = \gamma_{jk}^{-1}(\varepsilon)\gamma_{ij}^{-1}(\varepsilon)\, \frac{P_j(A_{ij})P_k(A_{ji})}{P_j(A_{ji})P_k(A_{jk})}$$

$$+ \gamma_{ij}^{-1}(\varepsilon)\gamma_{jk}^{-1}(\varepsilon)\, \frac{P_j(A_{kj})P_i(A_{jk})}{P_j(A_{jk})P_i(A_{kj})} \quad ,$$

and hence

$$D_\varepsilon \le \gamma_{jk}^{-1}(\varepsilon)\gamma_{ij}^{-1}(\varepsilon)\gamma_{jk} + \gamma_{ij}^{-1}(\varepsilon)\gamma_{jk}^{-1}(\varepsilon)\gamma_{ij} \ .$$

Next writing C as

$$C_\varepsilon = \gamma_{ij}^{-1}(\varepsilon) + \gamma_{jk}^{-1}(\varepsilon) + \gamma_{ik}^{-1}(\varepsilon)F_j(\varepsilon) \quad ,$$

where

$$F_j(\varepsilon) = \frac{P_i(A_{ij})P_i(A_{jk})P_i(A_{ki})P_k(A_{kj})P_k(A_{ji})P_k(A_{ik})}{P_i(A_{ji})P_i(A_{kj})P_i(A_{ik})P_k(A_{jk})P_k(A_{ij})P_k(A_{ik})} \ ,$$

it is seen that by setting i,j,k equal to the three cyclic permutations of $1,2,3$ we must have

$$F_1(\varepsilon)F_2(\varepsilon)F_3(\varepsilon) = 1 \ .$$

Hence i,j,k can be chosen such that

$$C_\varepsilon \le \gamma_{12}^{-1}(\varepsilon) + \gamma_{23}^{-1}(\varepsilon) + \gamma_{13}^{-1}(\varepsilon) \ .$$

Now it is easily verified that the maximal root of the equation (2.4), which is real and not smaller than 1, is an increasing and continuous function of both the coefficients C_ε and D_ε. Thus by Lemma 1 there is for every $\varepsilon > 0$ a 3-state RFA for which the error probability

$$P_e \le 1 - r^{-1} + \varepsilon \ , \tag{2.5}$$

where r is the maximal root of the equation

$$(1-\lambda)^3 - (1-\lambda)C_0 + D_0 = 0 \ , \tag{2.6}$$

with

$$C_0 = \gamma_{12}^{-1} + \gamma_{23}^{-1} + \gamma_{13}^{-1} \ ,$$

and

$$D_0 = \max\{\gamma_{12}^{-1} + \gamma_{23}^{-1}, \gamma_{12}^{-1} + \gamma_{13}^{-1}, \gamma_{23}^{-1} + \gamma_{13}^{-1}\} .$$

Clearly $(1/3)C_0 \leq (1/2)D_0$ and since $(1/2)D_0 \leq 1$ we must have $((1/3)C_0)^3 \leq ((1/2)D_0)^2$. Hence the maximal root of the cubic equation (2.6) is given by

$$r = 1 + 2 \, ((1/3)C_0)^{1/2} \cosh(1/3)\phi ,$$

where $\cosh \phi = 1/2 \, D_0 ((1/3)C_0)^{-3/2}$ and the bound (2.2) follows from (2.5). The simplified bound (2.3) can be obtained by increasing C_0 until $((1/3)C_0)^3 = ((1/2)D_0)^2$, the maximal root of (2.6) thus becoming

$$r = 1 + 2((1/2)D_0)^{1/3} .$$

3. COUNTEREXAMPLE TO TREE-CONJECTURE.

Consider a K-hypotheses problem and assume for simplicity that the support of each of the distributions P_k is same. With each RFA (A,Δ,d) we can now associate a graph Γ with vertices corresponding to states of the RFA and with an arc joining vertices i and j if and only if $p_{ij}p_{ji} \neq 0$. (This property does not depend on the hypothesis because of our assumption.)

Let $\varepsilon > 0$, C_ε^* be the class of all ε-optimal RFA, i.e., all m-state RFA (A,Δ,d) such that

$$P_e(A,\Delta,d) \leq \inf P_e(A,\Delta,d) + \varepsilon .$$

It has been conjectured by Cover [2] that for every $\varepsilon > 0$ the class C_ε^* always contains a RFA whose graph Γ is a tree. This is indeed true for $K = 2$ ([3]) and a plausible heuristic argument can be given for such a structure even for $K > 2$. Unfortunately, as we are going to show in this section, the conjecture is false already for $K = 3$. We do this by exhibiting a nontrivial 3-hypotheses problem and a 3-state RFA with a triangular graph Γ , which is strictly better than any 3-state RFA whose graph is a tree.

Let $X = \{1,2,3,4,5,6\}$, let p,q,r,s be positive numbers such that

$$2p + 2q + r + s = 1 ,$$

and

$$1 < \frac{r}{s} << \frac{p}{q} .$$

Define the three distributions p_1, p_2, p_3 as follows:

$P_k(x)$		x					
		1	2	3	4	5	6
	1	p	q	p	q	s	r
k	2	q	p	r	s	p	q
	3	s	r	q	p	q	p

Consider now a RFA (A, Δ, d) with the state space $S = \{1,2,3\}$, $d(i) = H_i$, $i \in S$ and the graph Γ the tree ①—②—③ . The matrix R for this RFA is the same as that on page 8 with $(i,j,k) = (1,2,3)$. Since by Lemma 1 the error probability is determined by the maximal eigenvalue ρ of R and ρ is always at least as large as the smallest row-sum, in order to minimize ρ we are forced to choose A as follows:

$$A_{12} = \{2\} , A_{21} = \{1\} , A_{23} = \{6\} , A_{32} = \{5\} .$$

The matrix R then becomes

$$R = \begin{bmatrix} 1 & , \frac{1}{p}\frac{\delta_{12}}{\delta_{21}} & , \frac{qr}{ps}\frac{\delta_{23}}{\delta_{21}}\frac{\delta_{12}}{\delta_{32}} \\ \frac{q}{p}\frac{\delta_{21}}{\delta_{12}} & , 1 & , \frac{q}{p}\frac{\delta_{23}}{\delta_{32}} \\ \frac{qs}{pr}\frac{\delta_{21}}{\delta_{23}}\frac{\delta_{32}}{\delta_{12}} & , \frac{q}{p}\frac{\delta_{32}}{\delta_{23}} & , 1 \end{bmatrix}$$

Writing its characteristic equation again as

$$(1-\lambda)^3 - (1-\lambda)C + D = 0$$

we have

$$C = 3(\frac{q}{p})^2 \ , \ D = (\frac{q}{p})^3 \ (\frac{s}{r} + \frac{r}{s}) \ . \qquad (3.1)$$

There are two other 3-state RFA whose graph Γ is a tree, one with the graph ②—①—③ and one with the graph ①—③—② . By the same reasoning as before we are forced tc choose

$$A_{12} = \{2\} \ , \ A_{21} = \{1\} \ , \ A_{13} = \{4\} \ , \ A_{31} = \{3\}$$

for the former, and

$$A_{13} = \{4\} \ , \ A_{31} = \{3\} \ , \ A_{23} = \{6\} \ , \ A_{32} = \{5\}$$

for the latter. The matrices R are

$$\begin{bmatrix} 1 & , \frac{q}{p} & , \frac{q}{p} \\ \frac{q}{p} & , 1 & , \frac{qs}{pr} \\ \frac{q}{p} & , \frac{qr}{ps} & , 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 \; , & \frac{qs}{pr} \; , & \frac{q}{p} \\ \frac{qr}{ps} \; , & 1 \; , & \frac{q}{p} \\ \frac{q}{p} \; , & \frac{q}{p} \; , & 1 \end{bmatrix}$$

respectively, where we omitted the δ's for the sake of simplicity. Hence the coefficient of their characteristic equations are again given by (3.1). Now consider a 3-state RFA (A, Δ, d) with

$$A_{12} = \{2\} \; , \; A_{21} = \{1\} \; , \; A_{13} = \{4\} \; , \; A_{31} = \{3\} \; , \; A_{23} = \{6\} \; , \; A_{32} = \{5\} \; ,$$

$\delta_{ij} = 1/2$ for all $i \neq j$ and $d(i) = H_i$, $i = 1,2,3$. The graph Γ of this RFA is a triangle. The matrix R is

$$\begin{bmatrix} 1 & \frac{q}{p}\,\frac{p+2s}{p+r+s} \; , & \frac{q}{p}\,\frac{p+2r}{p+r+s} \\ \frac{q}{p}\,\frac{p+2r}{p+r+s} \; , & 1 & \frac{q}{p}\,\frac{p+2s}{p+r+s} \\ \frac{q}{p}\,\frac{p+2s}{p+r+s} \; , & \frac{q}{p}\,\frac{p+2r}{p+r+s} \; , & 1 \end{bmatrix}$$

and the coefficients of its characteristic equation

$$C = 3\left(\frac{q}{p}\right)^2 \frac{p+2r}{p+r+s}\,\frac{p+2s}{p+r+s} = 3\left(\frac{q}{p}\right)^2 \left(1 - \left(\frac{r-s}{p+r+s}\right)^2\right) \; ,$$

and

$$D = \left(\frac{q}{p}\right)^3 \left[\left(\frac{p+2r}{p+r+s}\right)^3 + \left(\frac{p+2s}{p+r+s}\right)^3\right] \; .$$

Comparing these expressions with (3.1) we see that $C_{\text{triangle}} < C_{\text{tree}}$ and with r and s suitably chosen also $D_{\text{triangle}} < D_{\text{tree}}$. (Choose, for instance $r = 10^{-1}p$, $s = 10^{-3}p$. Then $D_{\text{tree}}/D_{\text{triangle}} = 10^4$). Since the maximal eigenvalue increases with both C and D we conclude from Lemma 1 that the best "tree" RFA has an error probability strictly larger than this "triangular" RFA.

4. MINIMAX THEOREM FOR FINITE-MEMORY PROBLEMS.

Let $\underline{\pi} = (\pi_1, \ldots, \pi_K)$ be a probability distribution on the set of hypotheses, let (A, Δ, d) be RFA, and let this time the error probability be

$$P_e(\underline{\pi}; (A,\Delta,d)) = \sum_{k=1}^{K} \pi_k P_e^{(k)}(A,\Delta,d) .$$

Looking now at the problem as a two-person zero-sum game, where the 1st player (Nature) chooses $\underline{\pi}$ and the 2nd player (Statistician) chooses (A,Δ,d) it is natural to ask whether

$$\inf_{(A,\Delta,d)} \sup_{\underline{\pi}} P_e(\underline{\pi}; (A,\Delta,d)) = \sup_{\pi} \inf_{(A,\Delta,d)} P_e(\underline{\pi}; (A,\Delta,d)) \tag{4.1}$$

Now if $K = 2$ then it is known [3] that

$$\inf P_e(\underline{\pi}, (A,\Delta,d)) = \frac{2(\pi_1 \pi_2 \gamma_{12}^{m-1})^{1/2} - 1}{\gamma_{12}^{m-1} - 1} \tag{4.2}$$

where γ_{12} is given by (2.1) . Hence

$$\sup \inf P_e = \frac{(\gamma_{12}^{m-1})^{1/2} - 1}{\gamma_{12}^{m-1} - 1} = 1 - (1 + \sqrt{\frac{1}{\gamma_{12}^{m-1}}})^{-1}$$

On the other hand by Lemma 1,

$$\inf \sup P_e = 1 - (\sup \rho(A,\Delta,d))^{-1}$$

and for $K = 2$ it is easily seen that

$$\rho(A,\Delta,d) = 1 + \left(\frac{\mu_1(d^{-1}(H_2))\mu_2(d^{-1}(H_1))}{\mu_1(d^{-1}(H_1))\mu_2(d^{-1}(H_2))}\right)^{1/2} .$$

However it has also been shown in [3] that

$$\sup \frac{\mu 1(d^{-1}(H_2))_{\mu 2}(d^{-1}(H_1))}{\mu_1(d^{-1}(H_1))\mu_2(d^{-1}(H_2))} = \frac{1}{\gamma_{12}^{m-1}}$$

and hence (4.1) is indeed true for $K = 2$.

Conjecture: (4.1) is also true for $K > 2$.

Comment: Since an analog of (4.2) for $K > 2$ is not available at present the above reasoning cannot be applied to prove the conjecture. However, since the number of hypotheses is finite (4.1) would follow if one could show that the set of all vectors

$$(P_e^{(1)}(A,\Delta,d),\ldots,P_e^{(K)}(A,\Delta,d)) ,$$

where (A,Δ,d) runs through all m-state RFA, is convex. This is indeed so for $K = 2$, unfortunately we have not been able to prove this even for $K = 3$.

APPENDIX A

A Formula For A Stationary Distribution Of A Finite Markov Chain.

Let $P = [p_{ij}]$ be an $m \times m$ stochastic matrix, let $g = (S,E)$ be an oriented graph with the set of vertices $S = \{1,\ldots,m\}$ and the set of arcs $E \subset S \times S$ defined by

$$(i,j) \in E \Longleftrightarrow i \neq j \quad \text{and} \quad p_{ij} > 0 \ .$$

Let $i \in S$ be a vertex of g. Then a vertex $j \in S$ such that $(i,j) \in E$ is called a successor of i. A sequence of vertices $(i_1, i_2, \ldots, i_n)$ such that each i_{k+1} is a successor of i_k, $k = 1,\ldots,n-1$, is called a path. If i_1 is also a successor of i_n the path $(i_1,\ldots,i_n)$ is called a cycle.

Consider now a subgraph $f = (S,F)$, where $F \subset E$ with the following properties.

1) each vertex $i \in S$ has at most one successor.

2) f has no cycles.

3) f is maximal, i.e., no further arcs can be added without violating 1) or 2).

We will call such a subgraph a confluence. Notice that each confluence has exactly one vertex with no successor. We will refer to this vertex as a sink.

With each confluence $f = (S,F)$ we associate a positive number

$$p(f) = \prod_{(i,j) \in F} p_{ij}$$

We now have the following theorem:

Theorem: Let P be a transition probability matrix of a homogeneous Markov chain, g be the graph defined above, let ϕ_i be the set of all confluences with sink $i \in S$.

If P has an invariant distribution $(\mu_1,\ldots,\mu_m) = (\mu_1,\ldots,\mu_m)P$ then

$$\mu_i = C \sum_{f \in \phi_i} p(f) \ , \quad i \in S \ , \tag{A.1}$$

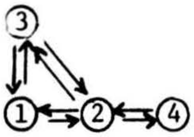
where $C > 0$ is a normalizing constant determined from $\mu_1 + \ldots + \mu_m = 1$.

Remark: Notice that the formula (A.1) gives μ_i as a sum of products of the off-diagonal entries of P, each product contains exactly $m - 1$ different entries and no two products contain the same set of p_{ij}'s. In this sense, the representation of μ_i is unique. Notice also, that if all off-diagonal entries of P are positive then μ_i is a sum of exactly m^{m-2} products. Thus, although the formula is certainly of theoretical interest, its application for computing the stationary distribution is likely to be limited to cases, where a majority of p_{ij}'s are zero.

Example: Let

$$P = \begin{bmatrix} .5 & .2 & .3 & 0 \\ .3 & .1 & .2 & .4 \\ .1 & .7 & .2 & 0 \\ 0 & .5 & 0 & .5 \end{bmatrix}$$

The graph g is



and the confluences together with the number $p(f)$ are as follows:

Sink i	Confluences $f \in \phi_i$			$\sum p(f)$
1	3 1 2 4 .015	3 1 2 4 .105	3 1 2 4 .010	.130
2	3 1 2 4 .010	3 1 2 4 .070	3 1 2 4 .105	.185
3	3 1 2 4 .045	3 1 2 4 .030	3 1 2 4 .020	.095
4	3 1 2 4 .008	3 1 2 4 .056	3 1 2 4 .084	.148
			Total:	.558

Hence the stationary distribution is $\underline{\mu} = \left(\frac{130}{558} , \frac{185}{558} , \frac{95}{558} , \frac{148}{558} \right)$.

Proof of the Theorem: We will show that (A.1) satisfy the equations

$$\mu_j = \sum_{i=1}^{m} \mu_i p_{ij} , \quad j=1,\ldots,m ,$$

or equivalently

$$\mu_j \sum_{\substack{i=1 \\ i\neq j}}^{m} p_{ji} = \sum_{\substack{i=1 \\ i\neq j}}^{m} \mu_i p_{ij} , \quad j=1,\ldots,m . \tag{A.2}$$

Let $h = (S,H)$ be an arbitrary subgraph of g , let

$$p(h) = \prod_{(i,j)\in H} p_{ij} ,$$

and let $h \pm (i,j)$ denote a subgraph obtained from h by adding or removing the arc (i,j). Next let

$$A_j = \{f + (j,i) : f\in \phi_j , i\in S - \{j\}\} ,$$

$$B_j = \{f + (i,j) : f\in \phi_i , i\in S - \{j\}\} .$$

If $\mu_1,\ldots,\mu_m$ is given by (A.1) then for any $j\in S$

$$\mu_j \sum_{\substack{i=1 \\ i\neq j}}^{m} p_{ij} = \sum_{h\in A_j} p(h) \quad \text{and}$$

$$\sum_{\substack{i=1 \\ i\neq j}}^{m} \mu_i p_{ij} = \sum_{h\in B_j} p(h) .$$

Thus (A.2) will follow if we show that $A_j = B_j$. To this end let

$$h\in A_j , \ h = f + (j,i) ,$$

let k be a vertex contained in the path $(i,\ldots,j)$ whose successor is j. If the arc (k,j) is removed then h becomes a conluence with sink k since (k,j) was an arc of confluence f and thus could not have any other successor than j. Hence h can be written as $f' + (k,j)$, $f' \in \phi_k$ and therefore $h \in B_j$. Conversly, if $h \in B_j$, say $h = f' + (k,j)$, then $f' \in \phi_k$ and by removing the arc (j,i) with i being a successor of j contained in the path $(j,\ldots,k)$ we conclude that $h - (j,i) \in \phi_j$. Hence $h \in A_j$ and the proof is complete.

APPENDIX B

A Generalization of a Lemma of Yakowitz ([5]).

Lemma: Consider K finite regular Markov chains with state spaces S_k, transition probabilities $[P_k(i\to j)]$, and stationary distributions $\underline{\mu}_k$, $k = 1,\dots,K$. Link these chains together by allowing transition between S_k and S_{k+1}, $k = 1,\dots,K-1$, via a pair of states $e_{k,k+1} \in S_k$, $e_{k+1,k} \in S_{k+1}$ with probabilities

$$P(e_{k,k+1} \to e_{k+1,k}) = \pi_{k,k+1} ,$$

$$P(e_{k+1,k} \to e_{k,k+1}) = \pi_{k+1,k} ,$$

and changing the original transition probabilities $P_k(e_{k,k+1} \to e_{k,k+1})$ and $P_{k+1}(e_{k+1,k} \to e_{k+1,k})$ accordingly.

If the new chain with state space $S = S_1 \cup \dots \cup S_K$ is regular then its stationary distribution $\underline{\mu}$ is given by

$$s \in S_k \Rightarrow \mu(s) = C\mu_k(s) \prod_{j=1}^{k-1} \pi_{j,j+1}\, \mu_j(e_{j,j+1}) \prod_{j=k+1}^{K} \pi_{j,j-1}\mu_j(e_{j,j-1}) , \qquad k = 1,\dots,K ,$$

where $C > 0$ is a normalizing constant.

Proof (by induction on K)

(i) Let $K = 2$. We have for the original two chains

$$s \in S_1 \Rightarrow \mu_1(s) = \sum_{r\in S_1} P_1(r\to s)\mu_1(s) ,$$

$$s \in S_2 \Rightarrow \mu_2(s) = \sum_{r \in S_2} P_2(r \to s)\mu_2(s) \ ,$$

and for the new chain $S_{12} = S_1 \cup S_2$,

$$s \in S_1, s \neq e_{12} \Rightarrow \mu_{12}(s) = \sum_{r \in S_1} P_1(r \to s)\mu_{12}(s) \ ,$$

$$s \in S_2, s \neq e_{21} \Rightarrow \mu_{12}(s) = \sum_{r \in S_2} P_2(r \to s)\mu_{12}(s) \ ,$$

$$\mu_{12}(e_{12}) = \sum_{r \in S_1 - \{e_{12}\}} P_1(r \to s)\mu_{12}(s) + [P(e_{12} \to e_{12}) - \pi_{12}]\mu_{12}(e_{12})$$

$$+\pi_{21}\mu_{12}(e_{21}) = \sum_{r \in S_1} P_1(r \to s)\mu_{12}(s)$$

since $\pi_{21}\mu_{12}(e_{21}) = \pi_{12}\mu_{12}(e_{12})$ by equating flows. Similarly

$$\mu_{12}(e_{21}) = \sum_{r \in S_2} P_2(r \to s)\mu_{12}(s) \quad .$$

Hence if $s \in S_1$ then $\mu_1(s)$ and $\mu_{12}(s)$ satisfy the same system of linear equations and consequently

$$s \in S_1 \Rightarrow \mu_{12}(s) = a_1\mu_1(s) \ ,$$

$$s \in S_2 \Rightarrow \mu_{12}(s) = a_2\mu_2(s) \ .$$

In particular

$$\mu_{12}(e_{12}) = a_1\mu_1(e_{12}) \ , \ \mu_{12}(e_{21}) = a_2\mu_2(e_{21})$$

and since $\pi_{12}\mu_{12}(e_{12}) = \pi_{21}\mu_{12}(e_{21})$ we must have

$$\frac{a_1}{a_2} = \frac{\pi_{21}}{\pi_{12}} \frac{\mu_2(e_{21})}{\mu_1(e_{12})} \quad .$$

Since $a_1 + a_2 = 1$ this implies $a_1 = C\pi_{21}\mu_2(e_{21})$, $a_2 = C\pi_{12}\mu_1(e_{12})$.

(ii) Let the lemma be true for $S = S_1 \cup \dots \cup S_{K-1}$ and form new chain $S \cup S_K$. Denoting $\underline{\mu}'$ the stat. distribution of S and $\underline{\mu}$ that of $S \cup S_K$ we have by part (i)

$$\begin{aligned} s \in S &\Rightarrow \mu(s) = C\,\pi_{K,K-1}\mu_K(e_{K,K-1})\mu'(s) \\ s \in S_K &\Rightarrow \mu(s) = C\,\pi_{K-1,K}\mu'(e_{K-1,K})\mu_K(s) \end{aligned} \tag{B.1}$$

By induction hypothesis

$$\mu'(e_{K-1,K}) = C'\mu_{K-1}(e_{K-1,K}) \prod_{j=1}^{K-2} \pi_{j,j+1}\,\mu_j(e_{j,j+1})$$

and if $s \in S_k \subset S$

$$\mu'(s) = C'\mu_k(s) \prod_{j=1}^{k-1} \pi_{j,j+1}\,\mu_j(e_{j,j+1}) \prod_{j=k+1}^{K-1} \pi_{j,j-1}\,\mu_j(e_{j,j-1}) \ .$$

Substitution into (B.1) gives the desired formula (with the proportionality constant CC').

REFERENCES

[1] Arbib, M. A., *Theories of Abstract Automata*, Prentice-Hall, (1969).

[2] Cover, T. M., personal communication (1974).

[3] Hellman, M. E., and Cover, T. M., "Learning with Finite Memory," *Ann. Math. Statist*. 41, (1970), pp. 765-782.

[4] Sagalowicz, D., "Hypothesis Testing with Finite Memory," Ph.D. thesis, Electrical Engr. Dept. Stanford University, (1970).

[5] Yakowitz, S., "Multiple Hypothesis Testing by Finite Memory Algorithms," *Ann. Statist*., 2, (1974), pp. 323-336.